DOCUMENT RESUME

ED 453 255 TM 032 792

AUTHOR Hansen, Joe B.; Schalock, Mark D.; McConney, Andrew; Rudd, Andy

TITLE Self Evaluation and Peer Observation of Early Career Teachers in a Standards-Based Context: Preliminary Results.

PUB DATE 2001-04-14

NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Academic Standards; *Beginning Teachers; Behavior Patterns; Elementary Education; *Observation; *Peer Evaluation; *Self Evaluation (Individuals); *Teacher Evaluation; Teaching Methods

IDENTIFIERS *Oregon

ABSTRACT

This study examined the relationships between teachers' self-assessment and their ratings by master teachers on key dimensions of proficiency as defined by Oregon's Continuing Teaching Licensure (CTL) requirements, and it explored differences between teacher behaviors and practices as a function of their teacher preparation programs. It also examined the properties of the observation instrument used in the study for the purpose of improving that instrument for possible use as a teacher assessment tool for the CTL process. The observation form, A Profile of Teaching in a Standards-Based Classroom, contains an observation section and a teacher rating form. Twelve master teachers with experience in teaching in Oregon received training in the use of the form and then observed sixty-five first-year, kindergarten through grade 5 teachers. This preliminary study, although far from conclusive, provided a considerable amount of useful information for refining the observation guide. Results show that teacher preparation institutions do have an effect on specific observable teacher practices in a standards-based framework. Data also suggest that higher levels of continuing professional development can lead to observable differences in teacher practices. The limited results obtained so far show promise for developing observation tools that are sensitive to identifying differences in teachers' practices and their preparation. (Contains 6 figures and 9 tables.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 453 255

Self Evaluation and Peer Observation of Early Career Teachers in a Standards-Based Context: preliminary results

Joe B. Hansen
Mark D.Schalock
Andrew McConney
Andy Rudd

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. Hansen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Teaching Research Division
Evaluation and Research Group Office
Western Oregon University

TM032792

Paper presented in the Division H symposium, *The Teacher Effectiveness Project – Phase-II: a mixed - method study of early career teachers in a standards-based schooling context*, at The Annual Meeting of the American Educational Research Association
Seattle, Washington, April 10-14

Contact Information:
Phone: (503) 838-8820
e-mail: hansenjb@wou.edu
fax: (503) 838-8150

BEST COPY AVAILABLE

Self Evaluation and Peer Observation of Early Career Teachers in a Standards-Based Context: preliminary results

Joe B. Hansen, Mark Schalock, Andrew McConney, Andy Rudd
Teaching Research Division
Western Oregon University

Background

This study examines the relationships between teachers' self-assessment and their ratings by master teachers on key dimensions of proficiency as defined by Oregon's Teaching Continuing Licensure (CTL) requirements, and explores differences between teacher behaviors and practices as a function of their teacher preparation program. It also examines the properties of the observation instrument used in the study for the purpose of improving that instrument for possible use as a teacher assessment tool for the continuing licensure process in Oregon. The context of this study is the *Teacher Effectiveness Study Part II* (TEPII) (McConney, Schalock, & Ayres, 2000; McConney & Schalock, 2001), a longitudinal study of teacher effectiveness. The even broader context is that of the recent requirement that all teachers in Oregon receiving their initial teaching license after 1999 must, if they want to continue teaching, meet the requirements for a Continuing Teaching License (CTL) by the end of their seventh year of teaching or face the loss of their teaching privileges.

A common core assessment system for CTL has been under development since fall, 1999. This system is being developed cooperatively by the teacher preparation institutions in Oregon under a Title-II grant through the Oregon University System. It is based on ten teaching proficiencies required by the Oregon Teacher Standards and Practices Commission, the teacher licensing agency for the State of Oregon. Under this new requirement, each teacher desiring to continue teaching after his or her seventh year will have to be certified by his or her teacher preparation institution that he/she has met the proficiencies shown in Figure 1.

Research Questions

The Teacher Effectiveness Project-phase-II (TEPII) asks the question: "*Does teacher preparation – in its design, structure, and character- make a difference in the practice, beliefs and characteristics of beginning teachers, and importantly, in the learning of their students?*"

This over-arching question has been unpacked to form three question groups; questions focused on comparative effects of teacher preparation programs on:

1. the *learning progress of students* taught by graduates of these students,
2. *teaching and learning to accepted standards*, and
3. *longitudinal teacher development* and the influence of context.

Figure 1. Advanced Proficiencies for Continuing Licensure

Advanced Teacher Proficiencies for Continuing Licensure

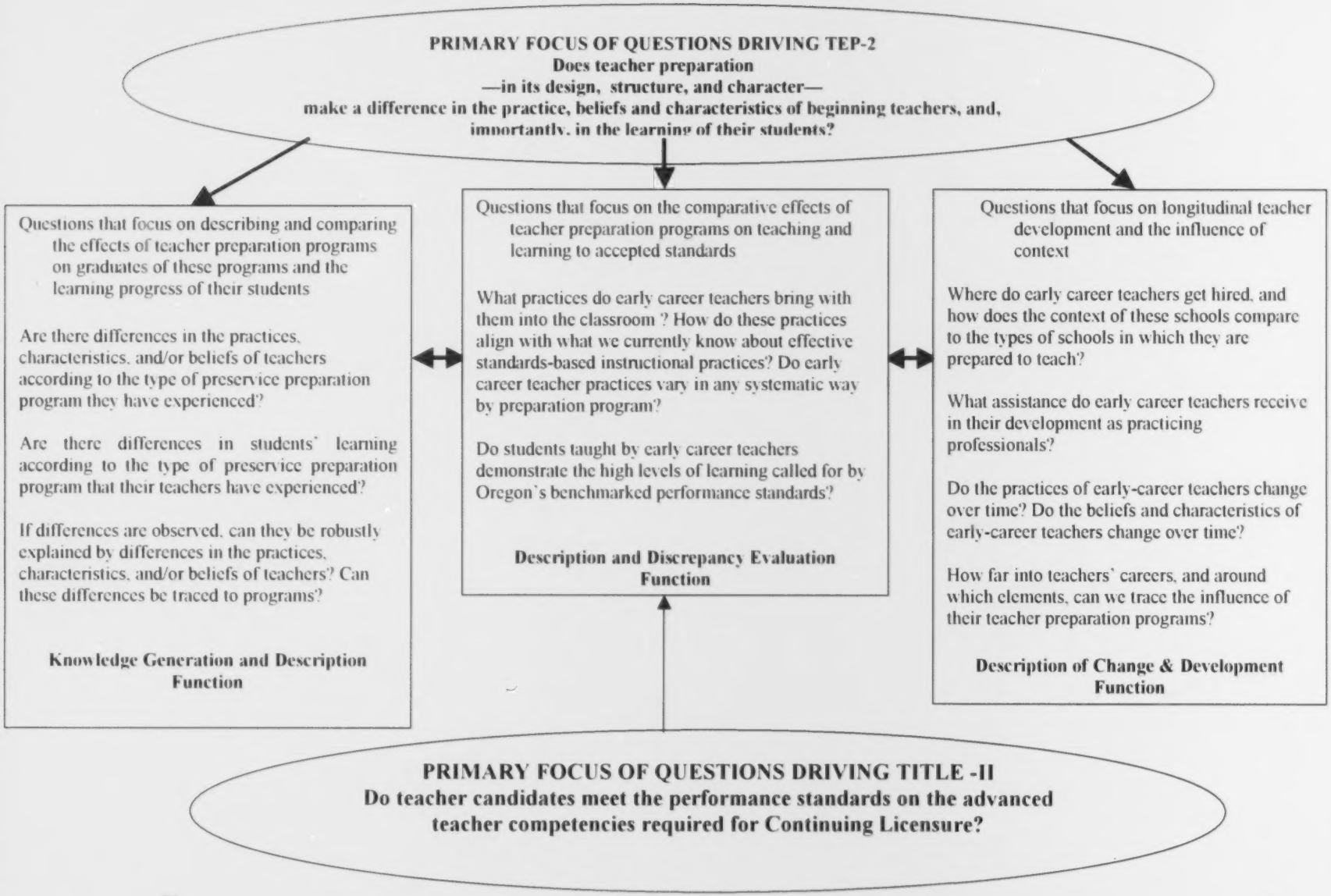
1. Assess knowledge and skills of students in relation to long-term content goals, State content standards, and district standards, and determine the knowledge and skills each student needs to accomplish them;
2. Design instructional plans that incorporate knowledge of students' developmental levels, interests, abilities, and learning accomplishments consistent with content standards and district standards;
3. Establish a classroom climate conducive to learning, e.g., positive classroom management, a safe and developmentally appropriate environment, efficient organization of time and materials, and effective transitions;
4. Implement instructional plans that employ knowledge of subject matter and use research-based educational practices that reflect how students learn, are sensitive to individual differences and diverse cultures, and encourage parent participation;
5. Evaluate student progress in learning, refine plans for instruction, and establish alternative learning options when necessary;
6. Document and report the progress of students in achieving content standards and district standards; and
7. Collaborate with parents, colleagues, and members of the community to provide internal and external assistance to students, and to their families;
8. Use emerging research on teaching, learning and school improvement to enhance practices;
9. Participate in designing, evaluating and improving opportunities for teaching and learning in an educational institution; and,
10. Collaborate with one's colleagues to enhance job performance and advance teaching as a profession.

Each of these questions has been further delineated into subsets of specific research questions addressed by the study. The schemata in figure 2 depicts all of the questions, their relationships to one another and to the continuing licensure process. This particular sub-study addresses, at least partially, four of these questions:

1. What practices do early career teachers bring with them into the classroom?
2. How do those practices align with what we currently know about effective standards-based instructional practices?
3. Do early career teachers' practices vary in any systematic way by preparation program?

A fourth question, not specifically identified in the TEPII schemata deals with the relationship between early career teachers' self perceptions of their proficiency and their observed practice in the classroom.

Figure 2. Conceptual Linkages Between TEP-2 and Title-II.


PRIMARY FOCUS OF QUESTIONS DRIVING TEP-2
Does teacher preparation
—in its design, structure, and character—
make a difference in the practice, beliefs and characteristics of beginning teachers, and,
importantly, in the learning of their students?
Questions that focus on describing and comparing the effects of teacher preparation programs on graduates of these programs and the learning progress of their students
Are there differences in the practices, characteristics, and/or beliefs of teachers according to the type of preservice preparation program they have experienced?
Are there differences in students' learning according to the type of preservice preparation program that their teachers have experienced?
If differences are observed, can they be robustly explained by differences in the practices, characteristics, and/or beliefs of teachers? Can these differences be traced to programs?
Knowledge Generation and Description Function
Questions that focus on the comparative effects of teacher preparation programs on teaching and learning to accepted standards
What practices do early career teachers bring with them into the classroom ? How do these practices align with what we currently know about effective standards-based instructional practices? Do early career teacher practices vary in any systematic way by preparation program?
Do students taught by early career teachers demonstrate the high levels of learning called for by Oregon's benchmarked performance standards?
Description and Discrepancy Evaluation Function
Questions that focus on longitudinal teacher development and the influence of context
Where do early career teachers get hired, and how does the context of these schools compare to the types of schools in which they are prepared to teach?
What assistance do early career teachers receive in their development as practicing professionals?
Do the practices of early-career teachers change over time? Do the beliefs and characteristics of early-career teachers change over time?
How far into teachers' careers, and around which elements, can we trace the influence of their teacher preparation programs?
Description of Change & Development Function
PRIMARY FOCUS OF QUESTIONS DRIVING TITLE -II
Do teacher candidates meet the performance standards on the advanced teacher competencies required for Continuing Licensure?

4. How do early career teachers' perceptions of their proficiency, as defined by the Oregon Continuing Teaching Licensure (CTL) requirements, relate to their effectiveness in applying standards-based instructional practices in the classroom?

The importance of this fourth question to the TEPII study and the CTL process is imbedded in the need for a common core assessment process for assessing teachers' readiness for continuing licensure. Such a system is essential to ensuring that teachers are held to high, uniform standards without compromising the character and individuality of the numerous teacher preparation programs throughout the state and thereby forcing all teachers to fit a single mold.

Instrumentation

Two of the measures under consideration for inclusion in the pool of assessment tools from which teacher preparation institutions my ultimately choose, are the *Teacher Self Assessment of Proficiencies for Continuing in Oregon* and an observation guide and rating form called *A Profile of Teaching in a Standards Based Classroom*.[1] These two instruments were developed by the TEP-II research team as components of a large battery of instruments and methods (McConney, Ayres, & Schalock, 2000.) The self-assessment contains two parts: one consisting of nine open-ended questions about the teacher's first year setting and induction into the profession and a second comprised of ten self-rating questions on the participant's level of proficiency as it relates to the Proficiencies For Continuing Licensure.[2]

The observation form, *A Profile of Teaching in a Standards-Based Classroom,* also has two components. Part 1, Identifying Specific Teaching Behaviors and Activities contains four subsections, each headed by a statement describing a standards-based educational practice: *1)communicating to students what is to be learned(Outcomes), 2)aligning and varying instructional activities, materials and procedures to support student outcomes, 3) assessing student progress and providing feedback, and 4)managing the classroom and engaging students.* Each of these subsections is comprised of a series of statements describing in simple non-inferential language, teacher classroom practices and behaviors that illustrate the concept represented by that subsection. The observer's task is to simply check whether that behavior occurred or not during an approximately ninety minute observation period. A space is provided after each statement for the observer to add clarifying notes regarding the observed occurrence.

Section 2 of the observation guide, the Summative Ratings, is a rating form on which the observer rates the teacher on each of seven different practices known to be associated with effective standards-based teaching, drawn from the research literature. These seven practices are intended to be congruent with specific behaviors and practices in the part-one checklist. The rating scale placed the teachers practices on a 0 to 5 point scale where 0 = not observed, 1 = poor, 2 = fair, 3 = good, 4 = very good and 5 = excellent.[3] It is important to point out that as a result of this first year's experience with this observation tool significant improvements were

[1] Earlier drafts of this instrument were known as the Standards Based Teaching Evaluation Protocol (STEP.)
[2] The self-assessment rating form document is too lengthy for inclusion in this paper, for further information about the form contact the author.
[3] It is important to note that the results described in this paper are based on version 1 of the observation protocol, which is no longer in use.

made to the scale and two teaching practices were added. Observers are encouraged to add comments on the back of the form and they usually do. The rating form is shown in figure 3.

Method

Procedure

Twelve master teachers with extensive experience teaching in Oregon were recruited to serve on a Regional Data Assistance Team (ReDAT) as part of the Core Assessment Development project under the Title-II grant (Hansen, 2000; Hansen, and Schalock, 2000). These teachers were given a full day of training on the TEP-II study, the CTL requirements and the use of the observation guide. They were selected in a manner consistent with the geographic distribution of the TEP-II participating teachers throughout the state. Each ReDAT member was assigned from four to five TEP-II teachers to observe. Each teacher had previously taken the self-assessment, but those results were not known to the ReDAT observers. Observations were scheduled by the ReDAT members to occur in the morning of a normal school day. Each observer was directed to obtain from the teacher a copy of his or her schedule for the day and structure the observation so that they could conduct two approximately equal length observations before lunch break. ReDAT members were instructed not to talk to the students during the observation periods. They were introduced by the teacher as "This is Mr./Mrs. Blank. She will be sitting in our classroom this morning, observing us as we do our work. Please don't pay any attention to her and don't talk to her. She will be very quiet and will not talk to you."

Sixty five first-year teachers in grades kindergarten through five were observed in this manner using the observation form.

Data Analysis

The Statistical Program for the Social Sciences (SPSS), version 9.0 was used to conduct all analyses. Microsoft Excel was also used to develop descriptive statistics and graphs for the self-assessment and observation data. Exploratory analyses consisting of bivariate correlations were conducted on each of the two data sets, self-assessment and observation For purposes of analysis, data from the two separate observations were averaged to provide observation scores on each item for each participant. Correlations were also computed self-assessment data were also analyzed to examine the relationships between the constructs represented by the self-assessment and observation. A principle components factor analyses was used to ascertain whether each instrument represented a separate, distinct construct or if items from one instrument might combine with items from the other to form a construct. Analyses were also conducted to look for differences in observed performance based on known differences among the teacher preparation programs from which the subjects graduated.

Observation Checklist Ratings. These data reveal a wide range of variability in the behaviors and practices observed. Observers reported that the checklist was easy to use and that it was relatively easy to determine whether a behavior or practice was present during the observation periods.

Observation Summative Ratings. Overall, these ratings were relatively high, with an overall mean of 3.84 on the five point scale, indicating the possibility of positive rater bias. These results are shown in Table 2. The "Total" column shows the overall means and standard deviations by rater across participants. The means are also displayed graphically in figure 4.

These data show considerable variability across observers. The relatively small standard deviations for some observers (numbers 4,5,6,8 and 12) could indicate some rater or sampling bias. This will be explored further through changing observers' assignments during year two.

Results

Descriptive Statistics

Observation Data. Section one of the observation form, the behavior checklist, was tallied and summed by participant to obtain numeric scores for each of its four sub-components: 1) Communicating to students what is to be learned (outcomes), 2) Aligning and varying instructional activities/materials to support student outcomes, 3) Assessing students and providing feedback, and 4) Managing the classroom and engaging students. The subscale scores for each participant were then averaged to create a composite score for each participant on each subscale. Similarly, the scores from the two separately obtained summative ratings from section two were combined to produce composite scores by item and overall for each participant. Descriptive statistics consisting of n, mean, range, minimum, maximum, standard deviation and variance were computed, for both sections of the observation instrument and the for the self-assessment rating scale, using SPSS. These data are displayed in Tables 1 a, 1b, and 1c.

Self-Assessment Data. Table 3a shows the frequency distribution for the self-assessment ratings. These are shown graphically in figure 5 a. Mean self-assessment ratings by proficiency are shown in table 3b and figure 5b. These data suggest a slight positive bias in the self-assessment ratings, especially when considering the fact that these are first year teachers.

Reliability

Estimates of reliability were obtained for the Self-Assessment and the Observation Guide using the covariance matrix method of SPSS. These analyses yielded alpha reliability coefficients of .9038 for the Self-Assessment scale and .9453 for the Observation Rating Scale. Although inter-rater reliability was recognized as important, it would have required placing at least two observers in the same classroom for a reasonably sized sample of teacher participants and this was not possible due to budgetary and time constraints. The TEPII research team still hopes to accomplish this before the project comes to an end in 2002.

Correlations

Bivariate, Pearson product-moment corelations were computed between all self-assessment items and the observation guide rating scale. These correlations, shown in table 4a include the inter-item correlations for each instrument. The correlation between the composite scores of the two instruments was low order and non-significant ($r = .131$, NS, $n = 65$.) Inter-item correlations for the Self-Assessment scale were in the range of $r = .282$ to .624, $p < .05$. For the Observation Rating Scale the inter-item correlations range from $r = .571$ to .836, $p < .05$. The inter-scale correlations were generally weak, showing little relationship between the items of one and items of the other. Correlations were also computed to examine the relationships between Section 1, the checklist, and Section 2, the rating scale, of the observation guide. These correlations were lower than anticipated and failed to show a strong and consistent relationship between actual observed practices and the summative ratings awarded teachers by their observers. These results are shown in table 4b.

Figure 3 Observation Rating Form

Section 2: Summative Ratings

Based on your observations, please provide a summative rating on each of the dimensions below by circling the appropriate number on the scale provided.

	Not Observed	Poor	Fair	Good	Very Good	Excellent
How well did this teacher communicate to students what was to be learned (outcomes of the lesson or activity?)	0	1	2	3	4	5
How well did this teacher align instruction and activities with communicated outcomes?	0	1	2	3	4	5
How well did this teacher vary activities and/or materials for students?	0	1	2	3	4	5
How well did this teacher assess/monitor students to adapt instruction?	0	1	2	3	4	5
How well did this teacher provide feedback to students about their work?	0	1	2	3	4	5
How well did this teacher engage students in learning activities?	0	1	2	3	4	5
How well did this teacher manage the classroom to maximize learning?	0	1	2	3	4	5

Comments:

Table 1a Observation Checklist Descriptive Statistics

	N	Range	Mini-mum	Maxi-mum	Mean	Std. Deviation
Communication	65	5.00	.00	5.00	3.5500	1.5419
Alignment	65	5.00	.00	5.00	3.8615	1.2915
Assessment/feedback	65	6.00	.00	6.00	4.0231	1.6475
Classroom Management	64	9.00	2.00	11.00	5.8281	1.6140
Valid N (listwise)	64					

Table 1b Observation Scale Section 2. Summative Rating, Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Communicates Outcomes	59	4.00	1.00	5.00	3.7831	.9884	.977
Alignment	59	4.00	1.00	5.00	4.0136	.9480	.899
Variety	61	3.50	1.50	5.00	3.8328	.9463	.896
Monitor and Adapt	60	4.00	1.00	5.00	3.6833	1.1460	1.313
Feedback	63	4.00	1.00	5.00	3.8778	1.1256	1.267
Engagement	65	4.00	1.00	5.00	3.8354	1.1452	1.311
Classroom management	64	4.00	1.00	5.00	3.9219	1.0846	1.176
Mean	65	4.00	1.00	5.00	3.7939	.9560	.914
Total	65	32.00	3.00	35.00	25.5292	7.6317	58.243
Valid N (listwise)	51						

Table 1c Self-Assessment Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Assessment	65	3.0	2.0	5.0	3.385	.878	.772
Planning	65	4.0	1.0	5.0	3.646	.891	.795
Classroom Management	65	4.0	2.0	6.0	4.231	.844	.712
Implementing Instruction	65	5.0	1.0	6.0	3.600	.844	.713
Collaboration	65	5.0	1.0	6.0	3.600	.981	.963
Evaluating Student Progress	65	3.0	2.0	5.0	3.831	.741	.549
Documenting Student Progress	65	4.0	1.0	5.0	3.523	1.032	1.066
Enhancing Practice	65	4.0	1.0	5.0	3.292	1.042	1.085
Professionalism	64	4.0	1.0	5.0	3.219	1.175	1.380
Collegiality	65	5.0	1.0	6.0	3.800	1.049	1.100
TOTAL	65	31.00	21.00	52.00	36.3077	7.0643	49.904
Valid N (listwise)	64						

Table 2 Observer Ratings by Item

Obs-erver	Statistic	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Obs. Total
1	Total	17	22	20	16	21	22	20	138
	Mean	4.25	4.40	4.00	3.20	4.20	4.40	4.00	4.06
	Std Dev	0.5	0.548	1.000	2.168	0.837	0.894	1.000	1.099
2	Total	52	54	44	52	52	51	49	354
	Mean	4.00	4.15	3.38	4.00	4.00	3.92	3.77	3.89
	Std Dev	1.581	1.345	1.710	1.354	1.291	1.320	1.235	1.386
3	Total	27	33	30	12	24	34	38	198
	Mean	2.45	3.00	2.73	1.09	2.18	3.09	3.45	2.57
	Std Dev	1.635	2.000	1.794	1.578	1.991	1.044	1.440	1.758
4	Total	20	22	23	20	20	23	23	151
	Mean	4.00	4.40	4.60	4.00	4.00	4.60	4.60	4.31
	Std Dev	1.000	0.894	0.894	1.414	1.225	0.894	0.894	0.993
5	Total	21	19	24	23	23	22	21	153
	Mean	4.20	4.75	4.80	4.60	4.60	4.40	4.20	4.50
	Std Dev	1.095	0.500	0.447	0.548	0.548	0.894	1.095	0.749
6	Total	39	40	36	39	43	40	41	278
	Mean	4.33	4.44	4.00	4.33	4.78	4.44	4.56	4.41
	Std Dev	0.707	0.527	0.707	0.500	0.441	0.527	0.527	0.586
7	Total	16	15	19	17	16	15	16	114
	Mean	3.20	3.00	3.80	3.40	3.20	3.00	3.20	3.26
	Std Dev	1.924	2.000	1.304	1.342	0.837	1.581	1.643	1.442
8	Total	20	22	18	20	25	25	24	154
	Mean	4.00	4.40	3.60	4.00	5.00	5.00	4.80	4.40
	Std Dev	0.000	0.548	0.548	0.000	0.000	0.000	0.447	0.604
9	Total	26	25	22	18	22	23	21	157
	Mean	3.25	3.13	2.75	2.25	2.75	2.88	2.63	2.80
	Std Dev	1.165	1.356	1.581	1.488	1.282	1.356	1.061	1.299
10	Total	34	33	25	40	46	47	52	277
	Mean	2.83	2.75	2.27	3.64	3.83	3.92	4.33	3.38
	Std Dev	1.850	2.137	2.005	0.809	1.528	1.240	0.492	1.645
11	Total	2	7	9	12	22	28	26	106
	Mean	0.20	0.70	0.90	1.20	2.20	2.80	3.25	1.56
	Std Dev	0.422	1.252	1.524	1.687	1.619	1.549	1.581	1.722
12	Total	39	36	42	41	43	38	40	279
	Mean	3.90	3.60	4.20	4.10	4.30	3.80	4.00	3.99
	Std Dev	0.994	1.430	0.789	0.738	0.675	0.789	0.816	0.909
13	Total	16	21	19	15	18	22	21	132
	Mean	2.00	2.63	2.38	1.88	2.25	2.75	2.63	2.36
	Std Dev	1.309	0.518	1.188	1.246	1.165	0.707	1.188	1.069

Figure 4 Mean Summative Ratings by Observer

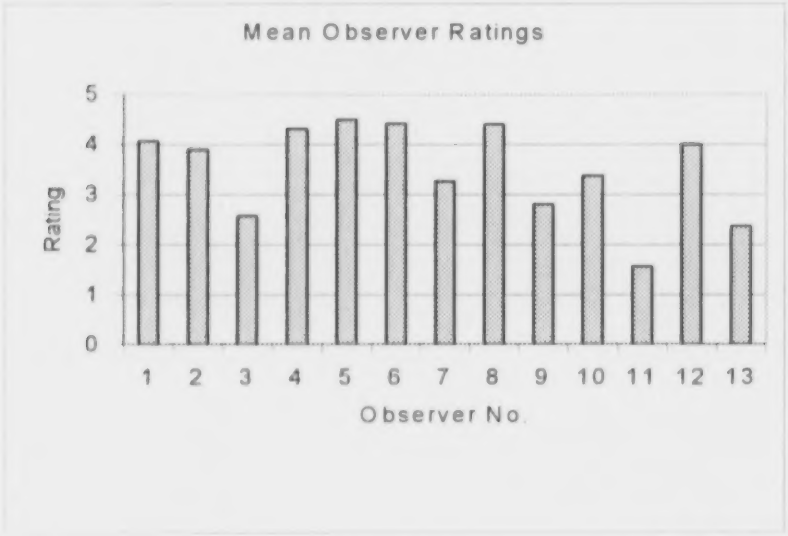


Table 3a Self Assessment Frequency Distribution

Proficiency Level	PR1	PR2	PR3	PR4	PR5	PR6	PR7	PR8	PR9	PR10	Total	Percent
Missing	0.0	2.0	0.0	1.0	2.0	0.0	3.0	2.0	5.0	1.0	16.0	2.63%
Developing	9.0	4.0	1.0	4.0	4.0	2.0	6.0	12.0	12.0	6.0	60.0	9.85%
Initial Prof	24.0	19.0	10.0	19.0	20.0	15.0	19.0	23.0	15.0	14.0	178.0	29.23%
Proficiency	23.0	25.0	26.0	31.0	26.0	33.0	23.0	15.0	20.0	24.0	246.0	40.39%
Expanding	5.0	11.0	22.0	5.0	7.0	11.0	10.0	9.0	8.0	16.0	104.0	17.08%
Advanced	0.0	0.0	2.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	5.0	0.82%
TOTAL	61.0	61.0	61.0	61.0	61.0	61.0	61.0	61.0	60.0	61.0	609.0	100.00%

Table 3b Self-Assessment Rating by Proficiency

	PR1	PR2	PR3	PR4	PR5	PR6	PR7	PR8	PR9	PR10	Total
n	61	61	61	61	61	61	61	61	60	61	609
sum	207	222	258	221	221	236	214	200	194	231	2204
mean	3.39	3.64	4.23	3.62	3.62	3.87	3.51	3.28	3.23	3.79	3.62
median	3	4	4	4	4	4	4	3	3	4	4.0
std dev	0.842	0.967	0.824	0.860	1.003	0.741	1.043	1.051	1.170	1.002	0.99228

Figure 5a, Self-Assessment Proficiency Self-Rating

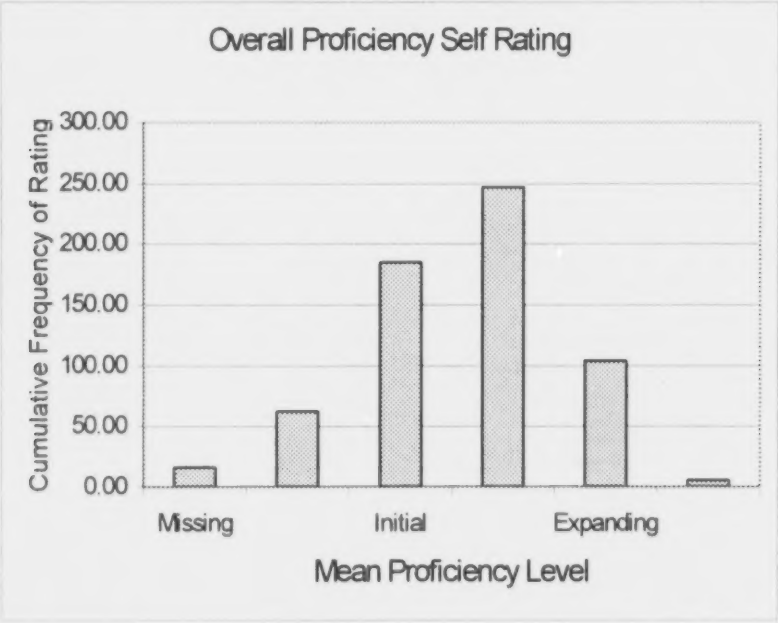
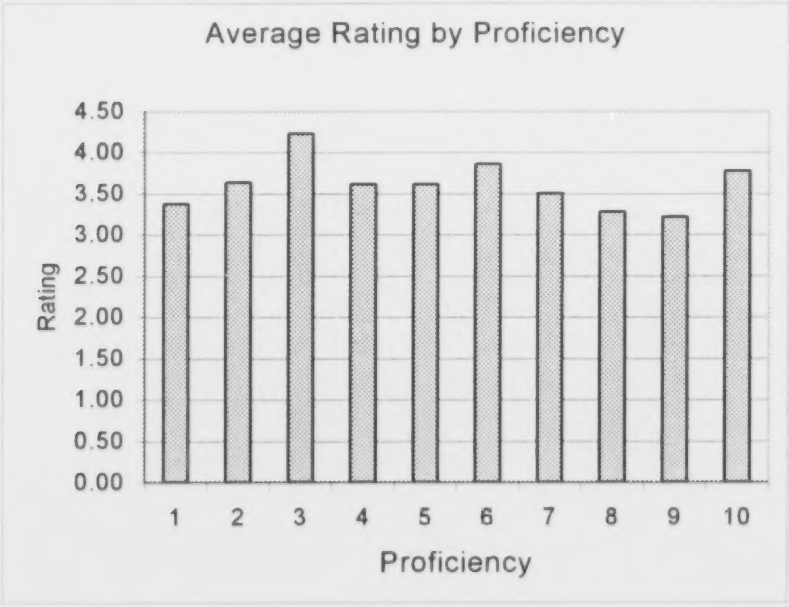


Figure 5b Self-Assessment Rating by Proficiency



Factor Structure

Observation form checklist items were summed by subsection for each of the four subsections; *communication, alignment, assessment feedback and classroom management.* This provided a subsection score for each participant. These scores were included with their corresponding participant ratings in a principle component factor analysis of the observation instrument. This analysis yielded two components or separate factors. The two factors represented the two different sections of the observation form, suggesting that they were not measuring the same set of behaviors. Separate, principle component factor analyses were conducted on each of the two separate rating scales, self assessment and observation, to ascertain whether they each represented single, coherent constructs or more than one construct. Each scale was found to form a single, coherent construct.

Differences by Institutional Characteristics

Study participants had received their teacher preparation training from 21 different institutions of higher education (IHE). Fifteen of these were within the state of Oregon and 6 were from outside the state. Institutions were sorted into four groupings: research oriented, regional public, private liberal arts, and private religious. Programs within the state of Oregon were classified into two groups, moderate and strong on the extent to which their programs were aligned with the Oregon State model for Standards Based Education adopted by the Oregon Department of Education. A one-way ANOVA was performed on the observation ratings to ascertain whether they varied systematically by which institution type they represented and if so, on which of the practices or behaviors they varied. These data are displayed in Table 5. Of the seven items representing standards-based teacher practices, significant differences among institutions were found for only two of the seven items, item 3, *Variety of instructional methods*: $F(3, 57) = 3.267$, $p < .05$ and item 7, *Manages classroom to maximize learning*: $F(3,59)\ F = 2.781$, $p < .05$. Two other items, item 2, *Alignment of Instruction with standards,* and 6, *Engages students in learning activities,* approached significance, as did the total score.

A one-way ANOVA on overall summative rating differences in observed practice between institutions on the extent (moderate or strong) to which they were standards-based, failed to reach significance at $p<.05$.

Participants also rated their preparation programs on other characteristics including program philosophy and the level of challenge they felt from their program. Each of these ratings was used to group participants into four groups and ANOVAs were performed examine whether such perceived program differences resulted in differences in observed teacher performance. These analyses produced non-significant between group results on individual items as well as on total observation ratings.

Table 4a Correlations between Self-Assessment Proficiency Ratings and Observation Guide Ratings

		Communicate Outcomes (Obs)	Alignment (Obs)	Variety (Obs)	Monitor and Adapt Inst. (Obs)	Feedback (Obs)	Engagement (Obs)	Classroom management (Obs)	Assessment (S-A)	Planning (S-A)	Classroom Management (S-A)	Implementing Instruction (S-A)	Collaboration (S-A)	Evaluating Student Progress (S-A)	Documenting Student Progress (S-A)	Enhancing Practice (S-A)	Professionalism (S-A)	Collegiality (S-A)
Communicate Outcomes	Pearson Correlation	1.000	.734**	.617**	.743**	.633**	.656**	.571**	.073	-.001	-.170	-.051	.021	-.070	-.034	-.115	-.068	.006
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.597	.993	.220	.714	.879	.613	.807	.408	.624	.965
	N	59	57	58	55	57	59	58	54	54	54	54	54	54	54	54	54	54
Alignment	Pearson Correlation	.734**	1.000	.671**	.709**	.664**	.687**	.673**	.236	.194	.372**	.050	.129	.114	.002	.118	.010	.182
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.089	.164	.006	.724	.356	.415	.988	.400	.944	.191
	N	57	59	57	55	57	59	58	53	53	53	53	53	53	53	53	53	53
Variety	Pearson Correlation	.617**	.671**	1.000	.701**	.588**	.606**	.592**	.148	.034	.231	.013	.057	.052	-.042	.070	-.066	-.075
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.277	.802	.086	.923	.676	.705	.759	.606	.628	.583
	N	58	57	61	57	59	61	60	56	56	56	56	56	56	56	56	56	56
Monitor and Adapt	Pearson Correlation	.743**	.709**	.701**	1.000	.786**	.747**	.723**	.123	.170	.233	.078	.194	.009	.043	.041	-.044	.088
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.375	.220	.090	.574	.160	.950	.758	.767	.753	.526
	N	55	55	57	60	60	60	59	54	54	54	54	54	54	54	54	54	54
Feedback	Pearson Correlation	.633**	.664**	.588**	.786**	1.000	.836**	.745**	.163	.158	.284*	.043	.125	-.025	.052	.006	-.217	.042
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.226	.240	.032	.752	.355	.856	.703	.965	.108	.757
	N	57	57	59	60	63	63	62	57	57	57	57	57	57	57	57	56	57
Engagement	Pearson Correlation	.656**	.687**	.606**	.747**	.836**	1.000	.816**	.137	.142	.288*	.057	.184	-.018	.161	.071	-.030	.012
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.302	.284	.027	.667	.163	.895	.223	.593	.824	.928
	N	59	59	61	60	63	65	64	59	59	59	59	59	59	59	59	58	59
Classroom management	Pearson Correlation	.571**	.673**	.592**	.723**	.745**	.816**	1.000	.244	.336**	.349**	.162	.283*	.091	.258	.165	.112	.196
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.065	.010	.007	.223	.046	.495	.050	.216	.408	.140
	N	58	58	60	59	62	64	64	58	58	58	58	58	58	58	58	57	58
Assessment	Pearson Correlation	.073	.236	.148	.123	.163	.137	.244	1.000	.465**	.537**	.521**	.503**	.542**	.504**	.497**	.400**	.512**
	Sig. (2-tailed)	.597	.089	.277	.375	.226	.302	.065		.000	.000	.000	.000	.000	.000	.000	.002	.000
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Planning	Pearson Correlation	-.001	.194	.034	.170	.158	.142	.336**	.465**	1.000	.522**	.549**	.363**	.606**	.449**	.289*	.282*	.481**
	Sig. (2-tailed)	.993	.164	.802	.220	.240	.284	.010	.000		.000	.000	.004	.000	.000	.025	.031	.000
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Classroom Management	Pearson Correlation	.170	.372**	.231	.233	.284*	.288*	.349**	.537**	.522**	1.000	.484**	.559**	.574**	.483**	.464**	.288*	.404**
	Sig. (2-tailed)	.220	.006	.086	.090	.032	.027	.007	.000	.000		.000	.000	.000	.000	.000	.027	.001
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Implementing Instruction	Pearson Correlation	-.051	.050	.013	.078	.043	.057	.162	.521**	.549**	.484**	1.000	.470**	.584**	.450**	.620**	.501**	.489**
	Sig. (2-tailed)	.714	.724	.923	.574	.752	.667	.223	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Collaboration	Pearson Correlation	.021	.129	.057	.194	.125	.184	.263*	.503**	.363**	.559**	.470**	1.000	.521**	.578**	.460**	.429**	.536**
	Sig. (2-tailed)	.879	.356	.676	.160	.355	.163	.046	.000	.004	.000	.000		.000	.000	.000	.001	.000
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Evaluating Student Progress	Pearson Correlation	-.070	.114	.052	.009	-.025	-.018	.091	.542**	.606**	.574**	.584**	.521**	1.000	.449**	.502**	.417**	.624**
	Sig. (2-tailed)	.613	.415	.705	.950	.856	.895	.495	.000	.000	.000	.000	.000		.000	.000	.001	.000
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Documenting Student Progress	Pearson Correlation	-.034	.002	-.042	.043	.052	.161	.258*	.504**	.449**	.483**	.450**	.578**	.449**	1.000	.598**	.499**	.417**
	Sig. (2-tailed)	.807	.988	.759	.758	.703	.223	.050	.000	.000	.000	.000	.000	.000		.000	.000	.001
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Enhancing Practice	Pearson Correlation	-.115	.118	.070	.041	.006	.071	.165	.497**	.289*	.464**	.620**	.460**	.502**	.598**	1.000	.561**	.547**
	Sig. (2-tailed)	.408	.400	.606	.767	.965	.593	.216	.000	.025	.000	.000	.000	.000	.000		.000	.000
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60
Professionalism	Pearson Correlation	-.068	.010	-.066	-.044	-.217	-.030	.112	.400**	.282*	.288*	.501**	.429**	.417**	.499**	.561**	1.000	.526**
	Sig. (2-tailed)	.624	.944	.628	.753	.108	.824	.408	.002	.031	.027	.000	.001	.001	.000	.000		.000
	N	54	53	56	54	56	58	57	59	59	59	59	59	59	59	59	59	59
Collegiality	Pearson Correlation	.006	.182	-.075	.088	.042	.012	.196	.512**	.481**	.404**	.489**	.536**	.624**	.417**	.547**	.526**	1.000
	Sig. (2-tailed)	.965	.191	.583	.526	.757	.928	.140	.000	.000	.001	.000	.000	.000	.001	.000	.000	
	N	54	53	56	54	57	59	58	60	60	60	60	60	60	60	60	59	60

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 4b Observation Checklist versus Rating Scale

		Communication (Check)	Alignment (Check)	Assessment/feedback (Check)	Classroom Management (Check)	Alignment (Check)	CommunicateOutcomes Rating	Variety Rating	Monitor / Adapt Rating	Feedback Rating	Engagement Rating	Classroom management Rating
Communication	Pearson Correlation	1.000	.757**	.704**	.671**	.366**	.370**	.316*	.285*	.276*	.285*	.317*
	N	65	65	65	64	58	58	60	59	61	63	62
Alignment	Pearson Correlation	.757**	1.000	.752**	.680**	.296*	.340**	.200	.174	.089	.082	.141
	N	65	65	65	64	58	58	60	59	61	63	62
Assessment/feedback	Pearson Correlation	.704**	.752**	1.000	.614**	.230	.369**	.292*	.246	.228	.151	.180
	N	65	65	65	64	58	58	60	59	61	63	62
Classroom Management	Pearson Correlation	.671**	.680**	.614**	1.000	.332*	.345**	.207	.241	.163	.107	.086
	N	64	64	64	64	57	57	59	58	60	62	61
Alignment	Pearson Correlation	.366**	.296*	.230	.332*	1.000	.734**	.671**	.709**	.664**	.687**	.673**
	N	58	58	58	57	59	57	57	55	57	59	58
CommunicateOutcomes	Pearson Correlation	.370**	.340**	.369**	.345**	.734**	1.000	.617**	.743**	.633**	.656**	.571**
	N	58	58	58	57	57	59	58	55	57	59	58
Variety	Pearson Correlation	.316*	.200	.292*	.207	.671**	.617**	1.000	.701**	.588**	.606**	.592**
	N	60	60	60	59	57	58	61	57	59	61	60
Monitor Adapt	Pearson Correlation	.285*	.174	.246	.241	.709**	.743**	.701**	1.000	.786**	.747**	.723**
	N	59	59	59	58	55	55	57	60	60	60	59
Feedback	Pearson Correlation	.276*	.089	.228	.163	.664**	.633**	.588**	.786**	1.000	.836**	.745**
	N	61	61	61	60	57	57	59	60	63	63	62
Engagement	Pearson Correlation	.285*	.082	.151	.107	.687**	.656**	.606**	.747**	.836**	1.000	.816**
	N	63	63	63	62	59	59	61	60	63	65	64
Classroom management	Pearson Correlation	.317*	.141	.180	.086	.673**	.571**	.592**	.723**	.745**	.816**	1.000
	N	62	62	62	61	58	58	60	59	62	64	64

** Correlation is significant at the 0.01 level (2-

* Correlation is significant at the 0.05 level (2-

Differences Related To School/District Context

Data were collected on a number of different context related variables to find out whether specific contextual factors made a difference in observed performance of first year teachers. One such variable was the amount of support provided for continuing professional development (CPD). Participants rated their schools on the amount of CPD by identifying which of five categories the school fell into 1) none, 2) very limited, 3) some-unfocused, 4) quite a bit – unfocused, and 5) substantial-focused. Due to small n sizes ANOVA across these groupings, was ruled out as being appropriate. However, when these categories were recoded into two - Low CPD support and High CPD support, by collapsing categories 1, 2 and 4 into the Low group and categories 4 and 5 into the High group a One-way ANOVA revealed significant between group differences on observation total rating, $F(1,60) = 7.609$, $p < .05$; observation mean rating, $F(1,60) = 4.145$, $p < .05$; and observation classroom management rating, $F(1,59) = 4.240$, $p < .05$; thus indicating that increased levels of CPD seemed to make a difference in observed teacher performance in classroom management skills, and overall performance. Another context related variable was participants' rating of their schools on the amount and type of mentoring support

BEST COPY AVAILABLE

provided. Participants rated their schools by placing the school in which they were employed in one of four categories: Limited Informal Help, Limited Formal Help, Lots of Informal Help, Lots of Formal Help. Cell sizes proved too small to justify a two factor (level by type of support) ANOVA. When these data were re-coded into two groups Limited Help and Lots of Help, by collapsing the informal v. formal dimension, they failed to produce a significant difference in a one-way ANOVA at $p<.05$.

Table 5 One Way ANOVA on Observation Summative Ratings by Institution

ANOVA

		Sum of Squares	df	Mean Square	F	Sig
Communicate Outcomes	Between Groups	5 356	3	1 785	1 914	138
	Within Groups	51 308	55	933		
	Total	56 663	58			
Alignment	Between Groups	6 627	3	2 209	2 670	056
	Within Groups	45 502	55	827		
	Total	52 129	58			
Variety	Between Groups	7 884	3	2 628	3 267	028
	Within Groups	45 850	57	804		
	Total	53 734	60			
Monitor and Adapt	Between Groups	4 712	3	1 571	1 209	315
	Within Groups	72 772	56	1 299		
	Total	77 483	59			
Feedback	Between Groups	2 508	3	836	717	546
	Within Groups	67 625	58	1 166		
	Total	70 134	61			
Engagement	Between Groups	6 873	3	2 291	1 995	124
	Within Groups	68 891	60	1 148		
	Total	75 764	63			
Classroom management	Between Groups	8 107	3	2 702	2 781	049
	Within Groups	57 329	59	972		
	Total	65 437	62			
Observation Total	Between Groups	247 013	3	82 338	2 529	069
	Within Groups	1529 933	47	32 552		
	Total	1776 945	50			
Observation Mean	Between Groups	5 463	3	1 821	2 423	075
	Within Groups	45 099	60	752		
	Total	50 562	63			

Discussion

This was a preliminary study based on data from the first year of a three-year longitudinal research project. As a preliminary study it had two broad purposes, 1) to examine the characteristics of the two instruments, the observation protocol and the self–assessment rating instrument, and 2) to begin to obtain, to the extent possible, answers to the research questions posed at the beginning of this paper. The research revealed a lot of useful information regarding the instruments, in particular the observation instrument. Much of that information has already been used in making improvements to the instruments for use this year and as the longitudinal study continues. It is expected that these improvements will make the instruments more useful to

the IHEs throughout Oregon who will be considering them as part of their tool kit for continuing licensure assessment. Some of these improvements will be discussed further.

Information bearing directly on answering the research questions was less abundant, due in part, to limitations in the instruments and small group sizes. Findings related to the research questions will also be discussed.

Instrument Characteristics and Implications for Improvement

Both instruments were found to be sufficiently reliable in terms of internal consistency to be useful for the research as well as for recommendation to the CTL assessment design team for use in assessing teacher readiness for continuing licensure. The subject of that assessment process is tangential to this paper, however it is important to note that the process is grounded in the beliefs that teacher performance is multi-dimensional and that multiple measures of teacher performance are essential for reasons of accuracy and fairness (Ferguson, D., 2001). Therefore, none of the assessment instruments or processes used for CTL assessment would be expected to stand alone.

The observation guide was revised in three substantive ways as a result of the first year's experience. First a new subsection was added to the observation checklist, focusing on the content of instruction. It became increasingly obvious to the TEPII research team that the practices and behaviors of teaching could not be completely separated from the ability of the teacher to impart meaningful content. Moreover, such data could be useful in explaining differences in student performance on content specific outcome measures.

A second revision to the observation instrument was that the observation rating scale was expanded from a five point to an eight point range and the points on the scale were given new labels corresponding to descriptive statements about the observed behavior or practice. These changes alone should make use of the instrument more accurate and less subjective. The new scale has three anchor points, one at each end and in the middle, with unambiguous descriptions of what each anchor point represents.

A third revision was the addition of two items to the rating scale, one dealing with how well the teacher promotes understanding and exploration of meaning across disciplines, and another with how well, holistically, the teacher promotes student interest in the learning content. At this writing, data collection for year two is nearly complete and early indications are that the revisions have resulted in increasing the variability in ratings and improving the overall quality of the data.

Limited Implications from Limited Results

At the beginning of this paper, four research questions to be addressed, at least partially by this study, were stated. Each of these will be addressed here, based on the results obtained.

Question 1. What practices do early career teachers bring with them into the classroom?

The observation instrument was designed to yield information about teachers' practices and classroom behaviors in seven broad areas known to be characteristic of standards-based instruction: 1) communication with students on what is to be learned (outcomes); 2) alignment of teaching and learning activities with and their support of communicated outcomes; 3) varying

instructional activities and materials to address learning needs of students; 4) assessing and monitoring student to adapt instruction; 5) providing feedback to students about their work; 6) engaging students in learning activities; and 7) managing the classroom to engage students in learning. The checklist portion of the instrument provided information on the occurrence of these behaviors and the rating scale provided a means of assessing how well the teachers performed in each of these areas. The checklist combined some of these categories to reduce the number of categories to four: 1) communicating outcomes, 2) aligning and varying activities and materials to support student outcomes, 3) assessing students and providing instruction, and 4) managing the classroom and engaging students. Part of the reason for combining these areas was practical, to reduce the number of pages in the instrument and make it less cumbersome to use. Additionally, it was felt after discussion by the research team, that in practice it was difficult to separate aligning instruction from varying the materials and methods and such separation was also difficult to discern between assessing students and providing feedback, especially when assessment is occurring informally, on the fly.

Observation Checklist data

The observation checklist was intended to provide descriptive data on the presence of certain essential standards based teaching practices and to provide the raw input that a ReDAT observer would use in forming the judgments required in performing the summative ratings. Given this expected relationship between the two parts, the corellations between checklist items and the items on the rating form were somewhat disappointing, and to some extent puzzling, in that they didn't conform to expectations in a clear way. In general, item pair correlations between the two sections were low, ranging from $r=.285$ ($p<.01$) to $r=.370$ ($p<.01$). Additionally, the strongest correlations between sections were between the communication items of the section and any items in the other scale. It may be that communication is more readily observable than the other behaviors and practices. These data are still being examined and a more thorough explanation is being sought.

Viewed through the TEP-2 lens, the observation data show a presence of standards-based teaching behaviors in these first year teachers, with considerable variability among them in terms in how strong that presence is.

Question 2. How do those practices align with what we currently know about effective standards-based instructional practices? The observation guide was developed through a lengthy process based on research on effective standards-based instruction and refined through extensive reviews by researchers and practitioners (McConney & Schalock, 2001.) Assuming the practices it addresses do then reflect effective standards-based instruction, these results are at best inconclusive. The orthogonality of the factor structure of the two sections of the guide and the lack of predictable relationships between the checklist items and the summative rating items requires further study. These results will be discussed in an upcoming forum with the ReDAT observers and the Title-2, 3.1 Design Team. It is hoped that insights based on first-hand experience in the field will aid the research team in understanding these seemingly anomalous results.

Question 3. Do early career teachers' practices vary in any systematic way by preparation program? The one way ANOVA to examine differences between preparation programs based on

their type of institution yielded significant results on specific items related to variety of instruction and classroom management. These results, although limited, are encouraging inasmuch as they suggest that at least certain differences that may be attributable to a teacher's IHE based, preparation program can be observed in the classroom. Analyses of data from year two should help to confirm or refute this finding.

Question 4. How do early career teachers' perceptions of their proficiency, as defined by the Oregon Continuing Teaching Licensure (CTL) requirements, relate to their effectiveness in applying standards-based instructional practices in the classroom? Some research suggests that people, or at least undergraduate college students, are ineffective in assessing their own competency level, when their self-ratings are compared with objectively obtained evaluation data, rating themselves higher than their objective test score (Kruger, J. & Dunning, D., 1999). The results of this study revealed some positive bias in both the self-ratings and the more objective observations, but also found a low correlation between self-ratings and observed performance. Therefore it is unclear from these results how useful self-perceptions can be in evaluating the performance of early career teachers. Year two data from the revised observation guide may be helpful in clarifying this issue. Until then we must conclude that little or no value can be derived form the self-assessment ratings for evaluative purposes. On the other hand, obervation data on effective standards-based teaching can only be fully validated by examining the relationship of observed effectiveness in concert with student achievement results. These analyses have begun and will be reported on later.

School Context. Although not specifically related to the questions posed for this study, data were also collected on numerous school context variables, most of which has not yet been analyzed. However, the analysis of one of these variables, the amount of continuing professional development support provided to new teachers in the field, did yield significant results, suggesting that receiving more support led to greater teacher effectiveness, specifically in the area of classroom management. These results are also encouraging and further support will be sought for this finding from second year data.

Summary and Implications

This preliminary study, while far from conclusive in any of the areas investigated, provided a considerable amount of useful information for refining the observation guide under consideration as one tool for CTL assessment in Oregon. In particular, encouraging results were obtained showing that teacher preparation institutions do have an effect on specific observable teacher practices, in a standards-based framework. The data also suggest that higher levels of continuing professional development support can lead to observable differences teacher practices. Data from the second year, now being analyzed will hopefully extend and clarify these findings.

The limited results obtained from this sub-study of the TEPII thus far show promise for developing observation tools that are sensitive to identifying differences in teachers' practices, traceable to their teacher preparation programs. These findings, if they can be confirmed and extended, will be helpful to Oregon's teacher preparation institutions as they fine-tune their support programs for continuing teacher licensure under new state requirements.

References

Ferguson, D. L. (2001)*A Core Teacher Licensure System for Continuing Licensure in Oregon*, working document prepared for the OQAT Title II initiative 3.1 Design Team.

Hansen, J.B. and Schalock, H.D. (2000) *Under the Microscope: the performance of first-year teachers in Oregon's standards-based schools*, presentation at the annual conference of the Education Trust, Washington, D.C., Nov.

Hansen, J. B. (2000) *using Peer Based Regional Assistance Teams to Assist Early Career Teachers Seeking continuing Licensure in a Research Context.* Paper presented at the Ninth Annual Evaluation Institute, Consortium for Research on Educational Accountability and Teacher Evaluation, San Jose, CA.

Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Peronality and Social Psychology*, 77, 6, 1121-1134.

McConney, A. and Schalock, M. (2001) *Developing an instrument for Observing Standards-Based Teaching*, Paper presented in the symposium The Teacher Effectiveness Project-Phase-II: a mixed-method study of early career teachers in a standards-based schooling context, at the American Educational Research Association Conference, Seattle, WA.

McConney, A., Schalock, M., and Ayres, R. (2000) *An Instrument for Observing Standards Based Teaching.* Paper presented at the Ninth Annual Evaluation Institute, Consortium for Research on Educational Accountability and Teacher Evaluation, San Jose, CA.